# Unified Semantic Parsing with Weak Supervision

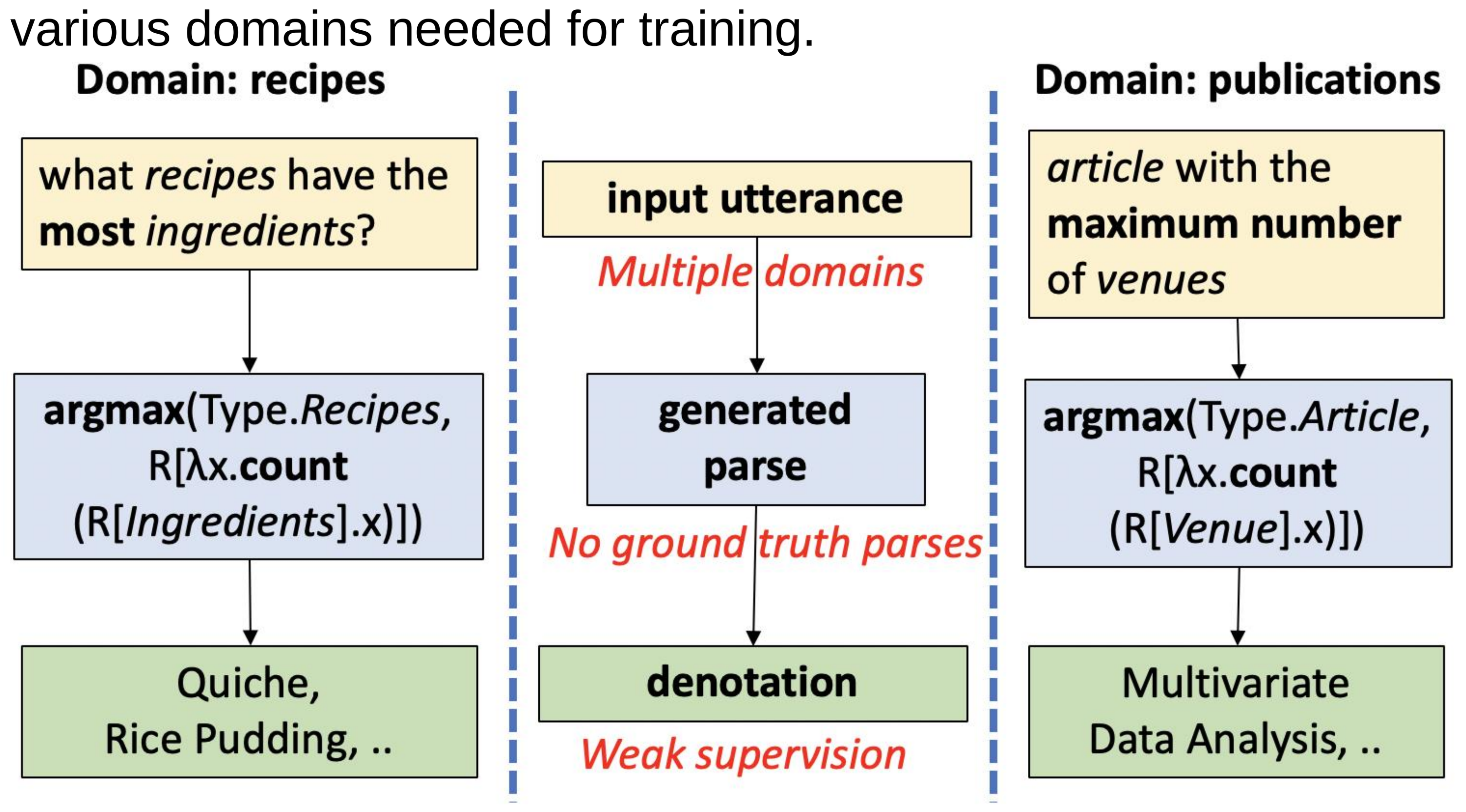

**Priyanka Agrawal*, Parag Jain, Ayushi Dalmia,**
**Abhishek Bansal, Ashish Mittal, Karthik S.**

**IBM Research AI**

*pagrawal.ml@gmail.com, {pajain34, adalmi08, abbansal, arakeshk, kartsank}@in.ibm.com*


## Problem

- Semantic parsing over multiple knowledge bases enables the parser to exploit structural similarities of programs across the multiple domains.
- Challenge lies in obtaining high-quality annotations of (utterance, program) pairs across various domains needed for training.




## Proposed Method

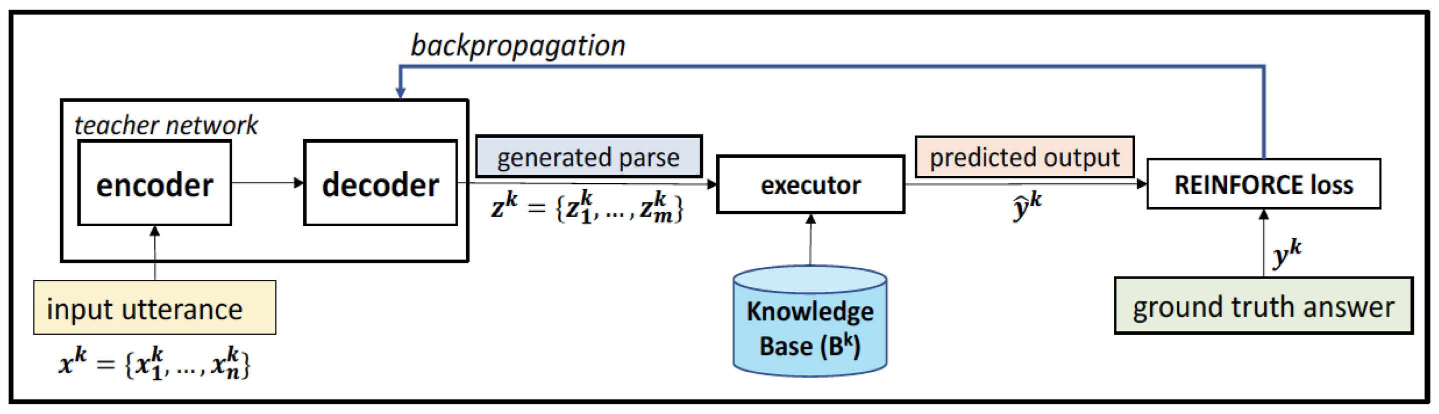
- Build a unified neural framework (DISTILLCOMBINED) to train **a single semantic parser** for multiple domains trained only with **weak supervision** i.e. using denotations in place of ground truth parse programs (Figure 1).
- Weakly supervised training is particularly arduous as the program search space grows exponentially in a multi-domain setting.
- We utilize multi-policy distillation mechanism:
  1. Train domain-specific semantic parsers (teachers) using weak supervision in the absence of the ground truth programs,

$$\sum_x \mathbb{E}_{P_\theta(z|x)}[R(x,z)] = \sum_x \sum_z P_\theta(z|x)R(x,z) \approx \sum_x \sum_{z\in B} P_\theta(z|x)[R(x,z)]$$

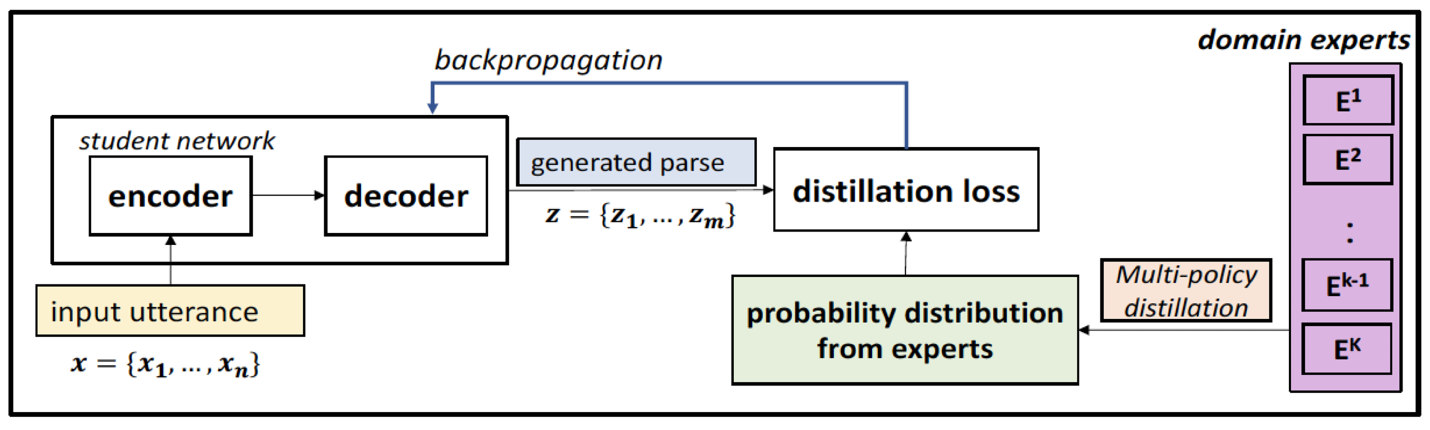
$\theta$ : policy parameters; B : output beam containing top scoring programs;
$P_\theta\,(z|x)$ : likelihood of parse z

  2. Train a single unified parser (student) from the domain specific policies obtained from these teachers T

$$L(\theta^S;\theta^1,...,\theta^K) = -\sum_{k=1}^{K}\sum_{i=1}^{|X^k|}\sum_{i=1}^{|m|}\sum_{v=1}^{|\mathcal{V}|} p_\theta^k(z_{ij}=v; x_i^k, z_{i\{1:j-1\}}) \log p_\theta^S(z_{ij}=v; x^k, z_{i\{1:j-1\}})$$




(a) Domain specific expert policy $E^k$




(b) Learning a unified student $S$ by distilling domain policies from experts $E^1, \cdots, E^K$

*Fig 1: Proposed Architecture Diagram*

## Results

### Baseline Models:

1. **WEAK-INDEPENDENT**: Set of weakly supervised semantic parsers, trained independently for each domain using REINFORCE algorithm.
2. **WEAK-COMBINED**: As per Herzig and Berant (2017), pool all the domains datasets into one and train a single semantic parser with weak supervision.
3. **DISTILL-INDEPENDENT: I**ndependent policy distillation for each domain.

**Dataset:** OVERNIGHT semantic parsing dataset (Wang et al., 2015)
Normalized to have reduced vocab size and search space)

| DOMAIN | WEAK-INDEPENDENT | WEAK-COMBINED | DISTILL-INDEPENDENT | DISTILL-COMBINED | SUPERVISED |
|---|---|---|---|---|---|
| BASKETBALL | 33.8 | 0.5 | 33.8 | **36.3** | 81.0 |
| BLOCKS | 27.6 | 0.8 | 36.8 | **37.1** | 52.8 |
| CALENDAR | **25.0** | 0.6 | 12.5 | 17.3 | 72.0 |
| HOUSING | 33.3 | 2.1 | 42.3 | **49.2** | 66.1 |
| PUBLICATIONS | 42.2 | 6.2 | 45.9 | **48.4** | 68.3 |
| RECIPES | 45.8 | 2.3 | 61.5 | **66.2** | 80.5 |
| RESTAURANTS | 41.3 | 2.1 | 40.9 | **45.2** | 73.5 |
| AVERAGE | 35.5 | 2.1 | 39.1 | **42.8** | 70.6 |

*Fig 2: Test denotation accuracy for each domain comparing our proposed method DISTILLCOMBINED with the three baselines. We also report the skyline SUPERVISED.*

- **Effect of Policy Distillation**: Policy distillation of individual expert policies result in an average percentage increase of ~ 10% in accuracy.
- **Performance of Unified Semantic Parsing framework:** DISTILL-COMBINED approach leads to an increased performance by ~ 20% in comparison to individual domain specific teachers
- **Effectiveness of Multi-Policy Distillation**: Due to weak signal strength and enlarged search space from multiple domains, WEAK-COMBINED model performs poorly across domains.

**Effect of Small Parallel Corpus**

Adding 10% parallel data brings an improvement of about 5 points, Increasing to only 30% givens improvement of about 11 points.

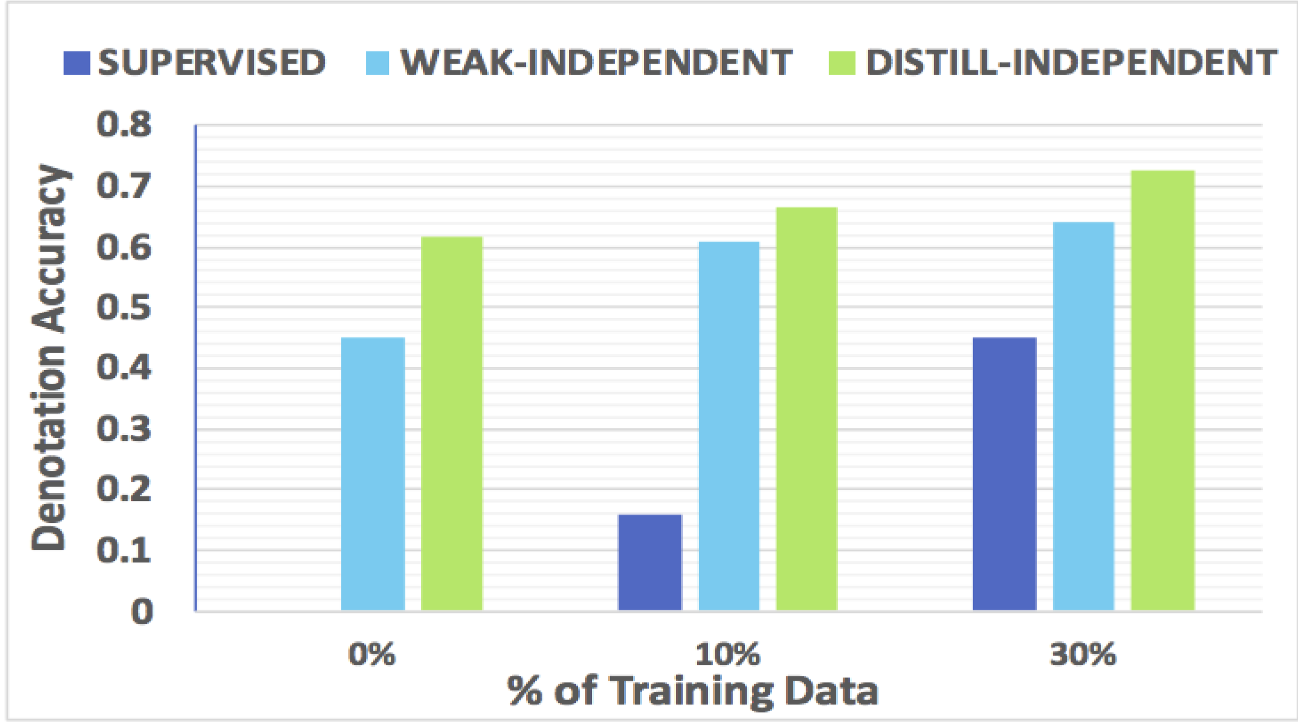



*Fig: 3 Effect of the fraction of training data on different models*

## References


1. Jonathan Herzig and Jonathan Berant. 2017. Neural semantic parsing over multiple knowledge-bases. In Association for Computational Linguistics
2. Yushi Wang, Jonathan Berant, and Percy Liang. 2015. Building a semantic parser overnight. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics.


* Priyanka has recently joined Booking.com. This work was done while she was a part of IBM Research